The ridge went north : Did the observer go as well? Corpus-driven investigation of fictive motion

Egorova, Ekaterina ; Boo, Gianluca ; Purves, Ross S

Abstract: Fictive motion ("The ridge went north") can refer to both dynamic (observer is moving) and static (observer is visually scanning) scenes. Using a corpus of alpine narratives, we extract fictive motion constructions and compare those representing static and dynamic scenes. According to our findings, some of the verbs appear exclusively in static (or dynamic) scenes, while others can be found in both and thus require broader context for the correct annotation. The results can be seen as a step towards the automatic identification of the role of geographic objects in text.

DOI: https://doi.org/10.21433/B3118354q145

Title
"The Ridge Went North": Did the Observer Go as Well? Corpus-driven Investigation of Fictive Motion

Journal
International Conference on GIScience Short Paper Proceedings, 1(1)

Authors
Egorova, Ekaterina
Boo, Gianluca
Purves, Ross S.

Publication Date
2016-01-01

DOI
10.21433/B3118354q145

Peer reviewed

"The Ridge Went North": Did the Observer Go as Well? Corpus-driven Investigation of Fictive Motion

E. Egorova[1,2], G. Boo[1], R.S. Purves[1]

[1]Department of Geography, University of Zurich, Zurich, Switzerland
Email: {ekaterina.egorova, gianluca.boo, ross.purves}@geo.uzh.ch

[2] University Priority Research Programme Language and Space (URPP SpuR), University of Zurich, Zurich, Switzerland

Abstract

Fictive motion ("The ridge went north") can refer to both dynamic (observer is moving) and static (observer is visually scanning) scenes. Using a corpus of alpine narratives, we extract fictive motion constructions and compare those representing static and dynamic scenes. According to our findings, some of the verbs appear exclusively in static (or dynamic) scenes, while others can be found in both and thus require broader context for the correct annotation. The results can be seen as a step towards the automatic identification of the role of geographic objects in text.

1. Introduction

Text corpora are increasingly being recognised as a potential source of rich geographic information. However, extracting information from text requires understanding of the ways in which language encodes space (Talmy 2000) and the scope of spatial information reproduced in certain discourse. Thus, route descriptions have been shown to go well beyond straightforward references to displacement through the prototypical *go* and *turn* (Allen 2000; Denis 1997; Tversky and Lee 1991) and include instructions of positioning and inspection (Allen 2000; Moncla *et al.* 2015), exercising caution or remembering a certain geographic object (Egorova *et al.* 2015), as well as topological information (Denis 1997). An important task in automatic route extraction from text is thus differentiating whether a geographic object is introduced as an element of the actual path or in some other context (e.g. description of a vista at some point of a path). One approach to making this distinction uses verb semantics, where a verb of perception signals description of the geographic object as part of a scene, while a verb of motion indicates movement of the observer with regard to that object. (Moncla *et al.* 2015).

From this perspective, fictive motion (FM) is of central importance. It depicts "the form, orientation, or location of a spatially extended object in terms of a path over the object's extent" (Talmy 2000) and reflects the conceptual primacy of named objects and their configuration in physical space (Matlock and Bergmann 2014; Matsumoto 1996). Crucially, it can take one of the two forms (Langacker 2010; Matsumoto 1996). The first (static) construes the scene as static, observable from a specific point in space, its conceptual roots lying in visual, or mental, scanning by an observer along the feature ("The path rises quickly near the top." (Langacker 2010)). The second (dynamic) encodes the actual motion of the observer, where the series of immediate fields of view along their path are construed as a single entity moving through space itself ("The path is rising quickly as we climb." (Langacker 2010)). Distinguishing between these forms is important, since one encodes the movement of the observer, while the other describes the configuration of an object in space.

In this study, we explore FM in a corpus of alpine texts addressing the following questions:

- Which verbs are used in FM constructions in our corpus?

- Can we identify differences between the way in which static and dynamic scenes are encoded in FM?

2. Corpus and Methods

Our corpus contains 1'484 texts (6'356'455 words) of the digitized Alpine Journal from 1968 to 2008 (Bubenhofer *et al.* 2015).

As seeds for FM extraction, we compiled a list of 125 nouns from climbing and mountaineering glossaries[1]. Although FM is associated with linearly extended entities, limiting the list to such features seemed inadequate, since it is the FM expression itself that construes the feature as linearly extended ("A table runs along the wall." (Matlock and Bergmann 2014)). Verbs of motion occur with both "structural" (e.g. *ridge, gulley, crack*) and "functional" (e.g. *route, pitch, line*) entities, we thus included both (Klippel 2003).

Using CQPWeb (Hardie 2012) to query a POS-tagged corpus by a lemma and a part of speech, we retrieved nouns in our list followed by a verb in the past or present tense. Among the candidate phrases we identified instances of FM (based on the verb semantics) and annotated those as static or dynamic. Markers for static types include verbs of perception ("we could see that the ridge rose in 4 steps"), locatives referring to the observer-centric field of view ("to the left, the face plunged"), the scale of the scene ("the range runs 1400 km") and, often, a verb's present tense for topological information. For the dynamic types, they include explicit references to the observer ("the ledge led us"), or implicit references to their movement (difficulty: "the pitch went easily"; time: "the ridge went forever"; past tense of verbs: "the crack ran to the top").

3. Results

Our initial query returned 6'530 phrases, of which 1'057 were FM; 655 of these were dynamic and 402 static. The range of verbs is large (81), reflecting the rich inventory language has for encoding FM. The specifics of corpus is reflected in verticality-related verbs (*rise, fall, drop, ascend, plunge, arise, mount, plummet, sink, shoot up*) and verbs semantically rich in the geometry of path (*curve, snake, wind, curl, roll, sneak, swing, weave, zigzag*). Verb frequency is Zipfian with 44 occurring only once or twice.

In the following, we focus on the top quintile represented by the 22 most frequent verbs (Figure 1). Few verbs appear exclusively with static or dynamic scenes, which in some cases seems to be determined by their semantics. Thus, *soar* has a strong connotation appealing to the verticality of a feature and the sensation of seeing it; *stretch* implies vast horizontal extension and is used for communicating general spatial knowledge ("the ridge stretches 8 km north"). Among the dynamic constructions, *bring* implies the presence and displacement of the second object. The majority of verbs, however, demonstrate semantic compatibility with both types of constructions.

[1]https://en.wikipedia.org/wiki/Glossary_of_climbing_terms, http://climber.org/data/glossary.html, http://www.ukclimbing.com/articles/page.php?id=33, http://www.summitpost.org/glossary-summit-peak-etc/173401, http://www.sierradescents.com/glossary, http://www.mountainzone.com/glossary_a_l.html

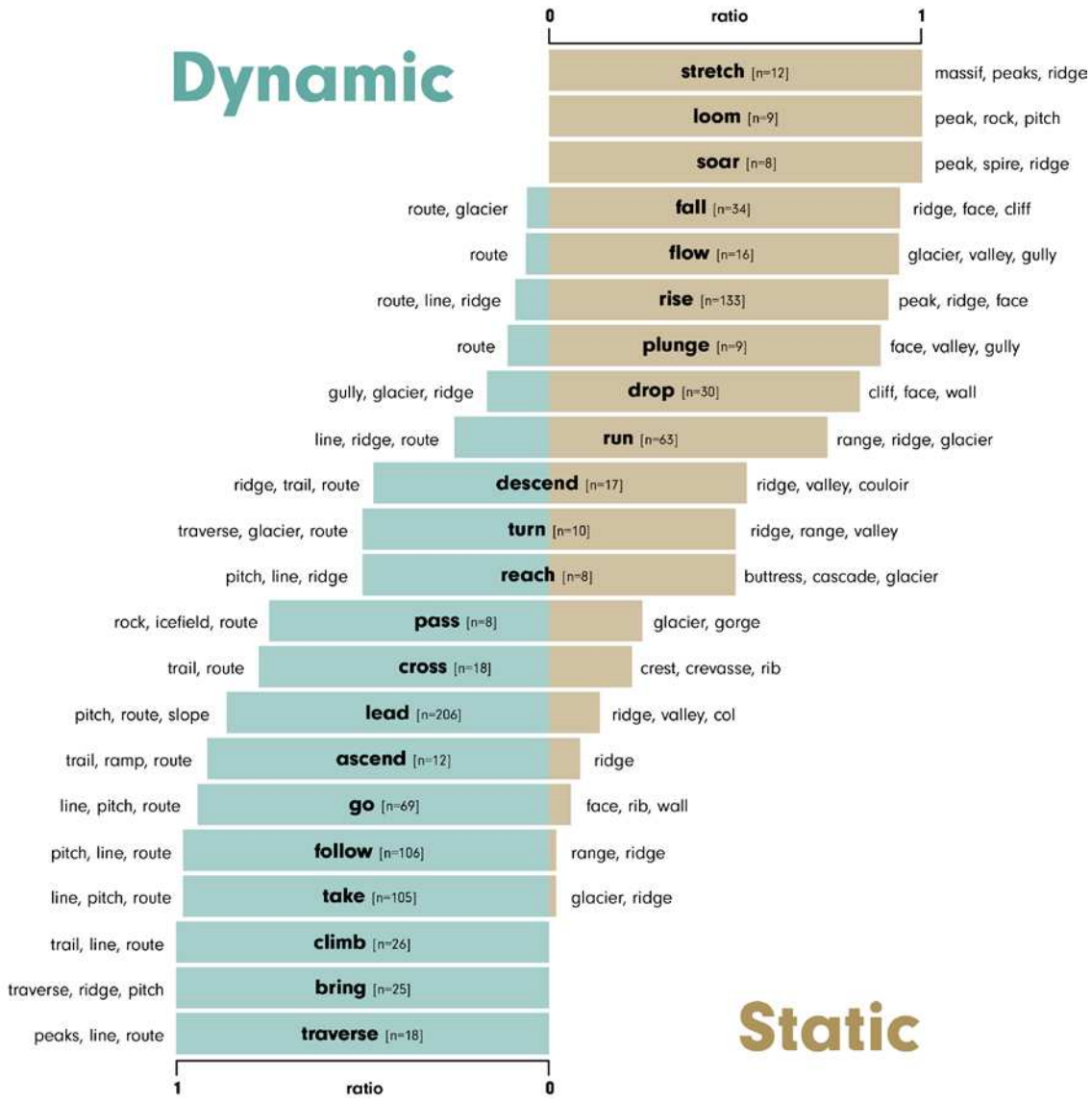

Figure 1. Ratio of static and dynamic uses of the 22 most frequent verbs and their three most frequent noun collocates.

When we explore frequent noun collocates with verbs the predominance of structural terms (e.g. *cliff, valley*) in static scenes becomes clear. In dynamic scenes, certain verbs clearly prefer functional collocates (e.g. *go, follow, take* with *line*, *pitch*, *route*), while others appear with both structural and functional terms. Thus, the same collocates (e.g. *ridge* and *descend*; *ridge* and *rise*) may encode both dynamic and static scenes and can only be annotated through surrounding context.

4. Conclusion and Outlook

Our corpus-driven exploration of FM in alpine narratives reveals a number of points important for further work. Using a set of nouns as seeds, we were able to extract a rich set of verbs used in FM. By developing a set of annotation rules linked to the context of collocates, it was possible to annotate these as either dynamic or static forms of FM. We found a tendency for some verbs to appear in dynamic (or static) scenes exclusively, for static scenes – to prefer structural entities,

which provides some indications as to possible features in automatic classification. However, some collocates remain ambiguous and contextual information (e.g. the markers we used in our annotation scheme) is required. In ongoing work we aim to evaluate the quality of our classification rules through inter-annotator agreement. Furthermore, we will systematically investigate the efficacy of the identified markers in the automatic classification of FM structures as part of our broader agenda for extracting semantically rich spatial information from text.

References

Allen GL, 2000, Principles and Practices for Communicating Route Knowledge. *Applied cognitive psychology*, 14(4):333–359.

Bubenhofer N, Volk M, Leuenberger F, Wüest D (eds), 2015, *Text+Berg-Korpus*. Institut für Computerlinguistik, Universität Zürich, Zürich, Switzerland.

Denis M, 1997, The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de psychologie cognitive*, 16(4):409–458.

Egorova E, Tenbrink T and Purves RS, 2015, Where Snow is a Landmark: Route Direction Elements in Alpine Contexts. In *Spatial Information Theory: 12th International Conference, COSIT 2015,* Santa Fe, NM, USA, 175–195.

Hardie A, 2012, CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3):380–409.

Klippel A, 2003, Wayfinding choremes. In *Spatial information theory: Foundations of geographic information science*. Springer, Berlin, Germany, 320–334.

Langacker RW, 2010, Dynamicity, Fictivity, and Scanning. In *Grounding Cognition: The Role of Perception and Action in Memory, Language and Thinking.* Cambridge University Press, Cambridge, UK, 164–197.

Matlock T and Bergmann T, 2014, Fictive Motion. In *Handbook of Cognitive Linguistics,* Walter de Gruyter, Berlin, Germany, 546–562.

Matsumoto Y, 1996, Subjective motion in English and Japanese. *Cognitive Linguistics*, 7(2):183–226.

Moncla L, Gaio M, Nogueras-Iso J and Mustière S, 2015, Reconstruction of Itineraries from Annotated Text with an Informed Spanning Tree Algorithm. *International Journal of Geographical Information Science*, 8816:1–24.

Talmy L, 2000, *Toward a Congitive Semantics*. The MIT press, Cambridge, MA, USA.

Tversky B and Lee PU, 1998, How Space Structures Language. *Spatial Cognition*. Springer, Berlin, Germany, 157–175.